

Étude de l'opération de c -jointure entre sous-décalages

par

Vivien DUCROS
ENS Paris-Saclay, Gif-sur-Yvette

encadré par

Enrico FORMENTI et Sandrine JULIA
CNRS, Université Côte d'Azur, Laboratoire I3S, France

Été 2022

Abstract

This internship report holds the results of the study of the c -join operation. The first result is the closure of regular languages by the c -join. The last one gives a characterization of the greatest set of forbidden words of the c -join on two subshifts, from their minimal sets of forbidden words.

Keywords : c -join, subshifts, bi-infinite words, cellular automata, dynamical systems (symbolic dynamics), formal languages, combinatorics on words, topology.

Résumé

Ce rapport de stage contient des résultats de l'étude de l'opération de c -jointure. Le premier résultat est la stabilité des langages rationnels par la c -jointure, le dernier donne une caractérisation du plus grand ensemble de mots interdits de la c -jointure de deux sous-décalages, à partir de leurs ensembles minimaux de mots interdits.

Mots-clés : c -jointure, sous-décalage, mots bi-infinis, automates cellulaires, systèmes dynamiques (dynamique symbolique), langages formels, combinatoire de mots, topologie.

Table des matières

| | |
|--|----------|
| Introduction | 2 |
| 1 Rappels et définitions | 3 |
| 1.1 Langages et automates | 3 |
| 1.2 Un peu de topologie | 4 |
| 1.3 Subshifts et ensembles de mots interdits | 5 |

| | | |
|----------|---|-----------|
| 2 | L'opération de c-jointure | 6 |
| 2.1 | Définition | 6 |
| 2.2 | Quelques propriétés | 7 |
| 3 | Premier résultat sur les langages rationnels | 7 |
| 3.1 | Stabilité des langages rationnels par la c -join : cas fini | 7 |
| 3.2 | En vue de la généralisation au cas infini | 9 |
| 4 | Résultats avec les langages facteurs et les mots interdits | 9 |
| 4.1 | La c -join et les langages facteurs | 10 |
| 4.2 | Caractérisation de l'ensemble maximal des mots interdits | 11 |
| | Conclusion et perspectives | 13 |
| | Annexes | 14 |
| A | Définition complète des préfixes, suffixes et facteurs | 14 |
| B | Correction de la construction de l'automate | 14 |
| C | Démonstration de l'associativité | 15 |

Introduction

Ce stage traite de l'étude de l'opération de c -jointure définie sur les subshifts (des langages de mots bi-infinis particuliers) dans un article de Enrico Formenti [FK07], encadrant de ce stage. Cette opération est utilisée pour déterminer l'ensemble maximal attracteur d'un automate cellulaire unidimensionnel.

Le stage s'est déroulé au laboratoire I3S de l'Université Côte d'Azur à Sophia-Antipolis. Il a duré 9 semaines pendant les mois de juin, juillet et août 2022.

Le sujet du stage était large et les axes de recherche étaient assez libres. Il était attendu initialement de déterminer des propriétés de l'opération de c -jointure, sans lien évident avec le contexte général. Cependant, au cours du stage, nous nous sommes orientés vers les effets de la c -join sur les langages facteurs et les ensembles de mots interdits, qui caractérisent les subshifts. Cela permet une meilleure compréhension du contexte de l'étude de cette opération et apporte des résultats théoriques complétant les bases déjà connues.

Certains termes sont en anglais dans ce rapport. Nous avons fait le choix de laisser ceux que nous utilisons à l'oral en anglais. Le mot *subshift* pourrait être traduit en français par *sous-décalage*. Pour ce qui est de la c -join, on écrit en français sans distinction *opération de c -jointure* ou *c -join*.

L'équipe dans laquelle j'ai travaillé était organisatrice de l'édition 2022 de l'École Jeunes Chercheuses et Chercheurs en Informatique-Mathématique pendant la première semaine de mon stage. J'ai pu assister aux conférences de chercheurs et de doctorants qui présentaient leur travaux. Mon encadrant de stage Enrico Formenti y a donné un cours d'introduction aux systèmes dynamiques et j'ai pu rencontrer mon autre encadrante Sandrine Julia. Tous deux sont enseignants-chercheurs permanents de l'équipe MDSC (Modèles Discrets pour les Systèmes Complexes).

J'ai participé à quelques activités organisées par l'association des doctorants : séminaires sur la carrière dans le monde académique après la thèse, afterworks, journée accrobranche un samedi.

1 Rappels et définitions

1.1 Langages et automates

Soit \mathcal{A} un ensemble fini dit *alphabet*. Les éléments de \mathcal{A} sont dits *lettres*. Sauf cas contraire, on aura $\mathcal{A} = \{0, 1\}$ dans tous les exemples. Un *mot fini* sur \mathcal{A} est une séquence finie d'éléments de \mathcal{A} . La *taille* $|u|$ d'un mot u est le nombre de lettres qui le composent. Le mot de longueur nulle est noté ε . On note \mathcal{A}^n l'ensemble des mots de taille $n \in \mathbb{N}$ pour un n fixé. On appelle *langage* un ensemble de mots. L'ensemble de tous les mots finis sur \mathcal{A} est noté $\mathcal{A}^* = \bigcup_{n \in \mathbb{N}} \mathcal{A}^n$. Un *mot infini* (à droite) (resp. *mot infini à gauche*) est un élément de $\mathcal{A}^{\mathbb{N}}$ (resp. $\mathcal{A}^{-\mathbb{N}}$).

Un *mot bi-infini* (pointé, au sens de [Car03]) est un élément de $\mathcal{A}^{\mathbb{Z}}$. On note $u = \cdots u_{-2}u_{-1} \cdot u_0u_1u_2 \cdots$.

Le *shift* (ou *décalage*, en français) est la fonction bijective $\sigma : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ telle que pour tout mot bi-infini w , on a : $\forall n \in \mathbb{Z}, \sigma(w)_n = w_{n+1}$. Pour deux mots bi-infinis u et v , on note $u \equiv v$ lorsqu'il existe $n \in \mathbb{N}$ tel que $u = \sigma^n(v)$. Notons que \equiv est une relation d'équivalence sur $\mathcal{A}^{\mathbb{Z}}$.

On s'autorisera lors de certaines preuves combinatoires à utiliser \mathcal{A}^{ζ} , l'ensemble quotient de $\mathcal{A}^{\mathbb{Z}}$ par la relation \equiv .

Définition 1 (Concaténation). *La concaténation $uv \in \mathcal{A}^*$ entre deux mots finis u et v est la juxtaposition de ces deux mots. On note $uv \in \mathcal{A}^{\mathbb{N}}$ la concaténation de $u \in \mathcal{A}^*$ et $v \in \mathcal{A}^{\mathbb{N}}$ et $uv \in \mathcal{A}^{-\mathbb{N}}$ lorsque $u \in \mathcal{A}^{-\mathbb{N}}$ et $v \in \mathcal{A}^*$.*

La concaténation pointée $u \cdot v \in \mathcal{A}^{\mathbb{Z}}$ de $u \in \mathcal{A}^{-\mathbb{N}}$ et $v \in \mathcal{A}^{\mathbb{N}}$ est définie par :

$$\forall k \in \mathbb{Z}, (u \cdot v)_k = \begin{cases} u_{k+1} & \text{si } k \leq -1 \\ v_k & \text{si } k \geq 0 \end{cases}$$

On étend la définition de concaténation aux langages de mots infinis pointés : $U \cdot V = \{u \cdot v \mid u \in U \wedge v \in V\}$.

Soit $(u, v) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$. La concaténation $uv \in \mathcal{A}^{\zeta}$ de u et v est définie par : $uv = \bigcup_{n \in \mathbb{Z}} \sigma^n(u \cdot v)$ (la classe d'équivalence de $u \cdot v$).

Exemple 1. Avec $u = \cdots u_{-2}u_{-1}u_0 \in \mathcal{A}^{-\mathbb{N}}$ et $v = v_0v_1v_2 \cdots \in \mathcal{A}^{\mathbb{N}}$ on a : $u \cdot v = \cdots u_{-2}u_{-1}u_0 \cdot v_0v_1v_2 \cdots$ et $uv = \cdots u_{-2}u_{-1}u_0v_0v_1v_2 \cdots$.

Définition 2 (Prefixes, suffixes, facteurs). *Soit $x \in \mathcal{A}^n$. Pour $0 \leq a \leq b \leq n$, le mot $x_{[a,b]} = x_a \cdots x_b$ est un facteur de x . Il est dit *préfixe* (resp. *suffixe*) lorsque $a = 0$ (resp. $b = n - 1$).*

Dans $\mathcal{A}^{\mathbb{N}}$, un *suffixe* est noté $x_{[a,+\infty[}$ pour $a \geq 0$.

Pour $x \in \mathcal{A}^{\mathbb{Z}}$, les facteurs de x sont :

- $x_{[a,b]}$, les facteurs finis
- $x_{]-\infty,b]}$, les préfixes
- $x_{[a,+\infty[}$, les suffixes
- $x_{]-\infty,+\infty[} = x$

où $a \leq b$ sont des entiers relatifs.

On notera $u \sqsubseteq x$ lorsque u est un facteur de x et $u \sqsubset x$ lorsque u est un facteur propre de x (ie. $u \sqsubseteq x$ et $u \neq x$).

On note $\text{Fact}(x)$ l'ensemble des facteurs de x , $\text{Pref}(x)$ l'ensemble de ses préfixes et $\text{Suff}(x)$ l'ensemble de ses suffixes.

Ces définitions s'étendent de manière naturelle aux langages comme suit :

- $\text{Fact}(L) = \bigcup_{x \in L} \text{Fact}(x)$
- $\text{Pref}(L) = \bigcup_{x \in L} \text{Pref}(x)$
- $\text{Suff}(L) = \bigcup_{x \in L} \text{Suff}(x)$

Définition 3. Un automate fini non-déterministe (resp. déterministe) est un quintuplet

$A = (\mathcal{A} \cup \{\varepsilon\}, Q, I, F, \delta)$ (resp. $A = (\mathcal{A}, Q, i, F, \delta)$) avec :

- \mathcal{A} un alphabet

- Q un ensemble fini d'états
- $I \subseteq Q$ l'ensemble des états initiaux (resp. $i \in Q$ l'état initial)
- $F \subseteq Q$ l'ensemble des états finaux
- $\delta \subseteq Q \times (\mathcal{A} \cup \{\varepsilon\}) \times Q$ l'ensemble des transitions
(resp. vérifiant $\forall (p, a, q, q') \in Q \times \mathcal{A} \times Q^2, (p, a, q) \in \delta \wedge (p, a, q') \in \delta \implies q = q'$)

On notera $p \xrightarrow[A]{a} q$ lorsque $(p, a, q) \in \delta$.

Définition 4. Une exécution d'un automate A est une suite finie de transitions $(p_i, u_i, q_i)_{i \in \llbracket 0, n-1 \rrbracket} \in \delta^n$ telle que : $\forall i \in \llbracket 0, n-1 \rrbracket, q_i = p_{i+1}$. Elle correspond à la lecture du mot $u = u_0 u_1 \cdots u_{n-1}$.

Elle est notée $p_0 \xrightarrow[A]{u_0} (q_0 = p_1) \xrightarrow[A]{u_1} \cdots \xrightarrow[A]{u_{n-1}} q_{n-1}$ ou $p_0 \xrightarrow[A]{u} q_{n-1}$.

Un mot est dit accepté par un automate lorsqu'il est lu lors d'une exécution de l'automate où p_0 est un état initial et q_{n-1} est un état final. Le langage reconnu par un automate est l'ensemble des mots acceptés par cet automate.

On note $\mathcal{L}(A)$ le langage reconnu par l'automate A .

Deux automates sont dits équivalents s'ils reconnaissent le même langage.

Définition 5. Un état est dit accessible s'il existe une exécution depuis un état initial vers cet état. Un état est dit co-accessible s'il existe une exécution depuis cet état vers un état final.

Un automate fini est dit émondé si tout état est à la fois accessible et co-accessible.

Proposition 1. Tout automate fini non-déterministe est équivalent à un automate fini déterministe.

Tout automate fini déterministe reconnaissant un langage non vide est équivalent à un automate fini déterministe émondé.

1.2 Un peu de topologie

Cette sous-partie introduit les notions topologiques utilisées pour définir les subshifts. Celles-ci sont données à titre indicatif pour comprendre le contexte des systèmes dynamiques symboliques, mais ne sont pas nécessaires à la compréhension globale de ce rapport.

Nous utilisons la topologie de Cantor sur les mots bi-infinis. Cette topologie est induite par la distance d suivante.

$$\forall x, y \in \mathcal{A}^{\mathbb{Z}} \quad d(x, y) = \begin{cases} 2^{-\min\{i \in \mathbb{N} \mid x_{-i} \neq y_{-i} \vee x_i \neq y_i\}} & \text{si } x \neq y \\ 0 & \text{sinon.} \end{cases}$$

Exemple 2. Soit $(x_n)_{n \in \mathbb{N}}$ la suite de mots bi-infinis telle que $\forall n \in \mathbb{N}, x_n = \sigma^n(0^{-\mathbb{N}} \cdot 1^{\mathbb{N}})$.

On a $x_0 = \cdots 00 \cdot 11 \cdots$, $x_1 = \cdots 001 \cdot 1 \cdots$ et ainsi de suite. La suite (x_n) a pour limite dans la topologie de Cantor le mot bi-infini $x = 1^{\mathbb{Z}}$.

La topologie de Cantor s'obtient aussi en prenant comme base les ensembles de configurations appelés *cylindres*.

Définition 6 (Cylindre). Soit $n \in \mathbb{N}^*$ et $u \in \mathcal{A}^n$, $m \in \mathbb{Z}$. Le cylindre de u à la position m est l'ensemble :

$$[u]_m = \{x \in \mathcal{A}^{\mathbb{Z}} \mid x_{[m, m+n-1]} = u\}.$$

En particulier, les boules ouvertes pour la distance d sont les cylindres centrés en 0, c'est à dire les $[u]_{-m}$ tel que $|u| = 2m + 1$.

On note \mathcal{B} l'ensemble des cylindres.

Les propriétés suivantes sont vérifiées :

- Tout cylindre est un ouvert et un fermé (ils sont aussi appelés *clopens*) ;

- Tout ouvert est réunion quelconque de cylindres (du fait que les cylindres sont une base) ;
- \mathcal{B} est un recouvrement de $\mathcal{A}^{\mathbb{Z}}$, c'est à dire $\mathcal{A}^{\mathbb{Z}} = \bigcup_{U \in \mathcal{B}} U$;
- L'intersection de deux cylindres est une union finie de cylindres.

Pour un ensemble $E \subseteq \mathcal{A}^{\mathbb{Z}}$, on note $\text{adh}(E)$ l'adhérence de E dans la topologie de Cantor (le plus petit fermé contenant E).

1.3 Subshifts et ensembles de mots interdits

Les subshifts sont les éléments de base de la dynamique symbolique, qui étudie des suites de symboles. Associée au codage, elle permet d'étudier les contraintes de stockage d'information dans les disques durs et les disques optiques. Différentes introductions à la dynamique symbolique permettent d'avoir un aperçu plus général du contexte : [BP97] [Kür03]. Des articles étudient certains cas particuliers de subshifts dont nous ne parlons pas ici : [Buz05] [Pav22].

Définition 7 (Subshift). *Un subshift Σ est un sous-ensemble de $\mathcal{A}^{\mathbb{Z}}$, fermé pour la topologie de Cantor et σ -invariant (c'est-à-dire tel que $\sigma(\Sigma) = \Sigma$).*

Les éléments d'un subshift sont appelés configurations.

Exemple 3. *Soit $\Sigma \subseteq \mathcal{A}^{\mathbb{Z}}$ un subshift contenant le mot $0^{-\mathbb{N}} \cdot 1^{\mathbb{N}}$. Alors, par σ -invariance il contient l'ensemble des mots de $\{\sigma^k(0^{-\mathbb{N}} \cdot 1^{\mathbb{N}}) \mid k \in \mathbb{Z}\}$. Puisque c'est un fermé, il contient aussi $0^{\mathbb{Z}}$ et $1^{\mathbb{Z}}$.*

Définition 8 (Langage facteur et ensemble de mots interdits). *Un subshift Σ est uniquement déterminé par son langage facteur $\mathcal{L} = \text{Fact}(\Sigma)$.*

Soit $D \subseteq \mathcal{A}^$. L'ensemble $\Sigma_D = \{x \in \mathcal{A}^{\mathbb{Z}} : \forall u \sqsubseteq x, u \notin D\}$ est le subshift d'ensemble de mots interdits D . Il est constitué de l'ensemble des mots bi-infinis qui n'ont pas de facteurs dans D .*

Définition 9. *Soit un subshift Σ . On définit les ensembles de mots interdits suivants :*

$$D^+(\Sigma) = \overline{\text{Fact}(\Sigma)}$$

$D^-(\Sigma) = \{x \in D^+(\Sigma) : \forall u \sqsubset x, u \notin D^+(\Sigma)\}$. *Cet ensemble est appelé collection of all first offender dans l'exercice 1.3.8 du livre de Lind et Marcus [LM95].*

On ne précisera pas forcément Σ lorsque c'est évident. On note parfois D_A^+ à la place de $D^+(\Sigma_A)$.

Proposition 2. *Soit un subshift Σ . On a les propriétés suivantes :*

1. $\Sigma_{D^+} = \Sigma$;
2. $\Sigma_{D^-} = \Sigma$;
3. *Soit D un ensemble de mots interdits tel que $\Sigma_D = \Sigma$. Alors $D \subseteq D^+$.
 $D^+(\Sigma)$ est donc le plus grand ensemble de mots interdits de Σ .*
4. *Soit D un ensemble de mots interdits tel que $\Sigma_D = \Sigma$. Alors $D \subseteq D^- \Rightarrow D = D^-$.
 $D^-(\Sigma)$ est donc un ensemble minimal de mots interdits de Σ .*
5. $D^+ = \mathcal{A}^* D^- \mathcal{A}^*$;
6. $\Sigma_{D_A^+} \cap \Sigma_{D_B^+} = \Sigma_{D_A^+ \cup D_B^+}$;
7. $\Sigma_{D_A^+} \cup \Sigma_{D_B^+} \subseteq \Sigma_{D_A^+ \cap D_B^+}$.

Exemple 4. *Soit $\Sigma = \{\sigma^k(0^{-\mathbb{N}} \cdot 1^{\mathbb{N}}) \mid k \in \mathbb{Z}\} \cup \{0^{\mathbb{Z}}, 1^{\mathbb{Z}}\}$ le subshift contenant uniquement les mots n'ayant que des 0 à gauche et des 1 à droite, auxquels on ajoute les mots ne comportant que des 0 ou que des 1. Alors $D^- = D^-(\Sigma) = \{10\}$. On a donc $\Sigma = \Sigma_{\{10\}}$.*

D^- n'est pas le plus petit ensemble de mots interdits, car avec $D = \{100, 101\}$, on a $\Sigma_D = \Sigma$, or $D^- \not\subseteq D$.

2 L'opération de c -jointure

La c -jointure a été définie initialement sur les subshifts dans l'article de Formenti et Kůrka [FK07]. Cette définition a été retravaillée pour être adaptée aux raisonnements et aux démonstrations présentés dans ce rapport.

2.1 Définition

Voici une présentation de la définition originale, puis des définitions que nous avons utilisées au cours de mon stage. Il s'est avéré utile de raffiner la définition au fur et à mesure pour qu'elle puisse s'adapter à nos raisonnements.

Définition 10 (Définition originale de la c -join, [FK07]). *Soit $c \in \mathbb{N}$ et deux subshifts Σ_A et Σ_B . La c -join $\Sigma_A \overset{c}{\vee} \Sigma_B$ de ces subshifts est l'ensemble des configurations $x \in \mathcal{A}^{\mathbb{Z}}$ telles que : soit $x \in \Sigma_A \cup \Sigma_B$, soit il existe deux entiers a et b tels que $b - a \geq c$, $x_{]-\infty, b-1]} \in \text{Pref}(\Sigma_A)$ et $x_{[a, +\infty[} \in \text{Suff}(\Sigma_B)$.*

La c -join de deux subshifts est donc l'union de ces deux subshifts, à laquelle on ajoute les mots créés par superposition d'une portion de taille au moins c de mots du premier subshift à gauche et du second subshift à droite.

Nous avons remarqué au bout de deux semaines que la condition $b - a \geq c$ est superflue. Le lecteur pourra se convaincre, au vu des définitions suivantes, qu'il suffit de la condition $b - a = c$ dans la définition. Cela nous permettra de faciliter certains raisonnements.

La définition suivante est appliquée aux subshifts. Par souci de clarté, on utilise la notation des mots dans $\mathcal{A}^{\mathbb{Z}}$.

Définition 11. *Soient Σ_A et Σ_B deux subshifts de $\mathcal{A}^{\mathbb{Z}}$ et c un entier positif. La c -jointure de ces subshifts est définie ainsi (voir figure 1) :*

$$\Sigma_A \overset{c}{\vee} \Sigma_B = \Sigma_A \cup \Sigma_B \cup \{uvw : w \in \mathcal{A}^c \wedge uw \in \text{Pref}(\Sigma_A) \wedge vw \in \text{Suff}(\Sigma_B)\}$$

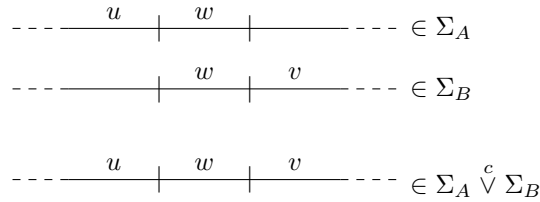


FIGURE 1 – Configurations créées par superposition

Exemple 5. *Soit $\Sigma_A = \Sigma_{\{111\}}$, le subshift des mots ne comportant pas trois 1 consécutifs et $\Sigma_B = \Sigma_{\{10\}}$ le subshift des mots où une occurrence d'un 1 ne peut être suivie que d'un autre 1.*

Alors $\Sigma_A \overset{1}{\vee} \Sigma_B = \Sigma_{\{1110\}}$, le subshift des mots où toute occurrence de trois 1 consécutifs est suivie d'un autre 1. La présence du facteur 111 marque le "passage" de la superposition, et indique que les lettres à sa droite sont celles d'un mot de Σ_B .

Nous avons aussi adapté la définition aux langages de mots finis pour le début du stage.

Définition 12. *Soient L_A et L_B deux langages de \mathcal{A}^* et c un entier positif. La c -jointure de ces langages est définie ainsi (voir figure 2) :*

$$L_A \overset{c}{\vee} L_B = L_A \cup L_B \cup \{uvw : w \in \mathcal{A}^c \wedge uw \in \text{Pref}(L_A) \wedge vw \in \text{Suff}(L_B)\}$$

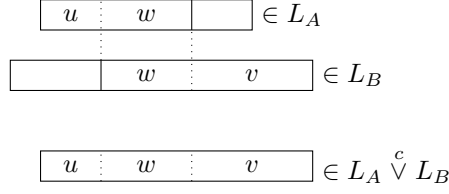


FIGURE 2 – Mots créés par superposition

Nous avons été amenés à utiliser la notation suivante pour les mots créés par superposition :

$$J_c = \{uvw : w \in \mathcal{A}^c, uw \in \text{Pref}(\Sigma_A) \text{ et } wv \in \text{Suff}(\Sigma_B)\}.$$

La c -join peut donc s'écrire $\Sigma_A \overset{c}{\vee} \Sigma_B = \Sigma_A \cup \Sigma_B \cup J_c$.

Cette notation est abusive car les subshifts Σ_A et Σ_B n'y sont pas précisés. On réserve donc son usage lorsqu'il n'y aura pas de confusion possible.

Une définition a été ajoutée dans la partie 4.2 pour démontrer le dernier théorème. Pour ne pas surcharger le lecteur, nous avons fait le choix de ne pas l'introduire ici. Il n'est pas nécessaire de la lire maintenant, il vaut mieux se familiariser avec les définitions précédentes d'abord.

2.2 Quelques propriétés

L'étude de premières propriétés plus ou moins immédiates a permis de nous familiariser avec l'opération de c -join.

Proposition 3 (Énoncée dans [FK07]). *Pour tout $c \in \mathbb{N}$ et toute paire de subshifts Σ_A et Σ_B , $\Sigma_A \overset{c}{\vee} \Sigma_B$ est un sous-décalage.*

Proposition 4. *Pour $x \in \mathcal{A}^{\mathbb{Z}}$, on note \tilde{x} le mot miroir de x , tel que $\forall n \in \mathbb{Z}, \tilde{x}_n = x_{-n}$. On note aussi $\tilde{E} = \{\tilde{x} \mid x \in E\}$. Soit $c \in \mathbb{N}$ et deux subshifts Σ_A et Σ_B . Alors : $\widetilde{\Sigma_A \overset{c}{\vee} \Sigma_B} = \tilde{\Sigma}_B \overset{c}{\vee} \tilde{\Sigma}_A$.*

Proposition 5. $\forall c \in \mathbb{N}, J_{c+1} \subseteq J_c$

Remarque 1. *La proposition 3 nous apprend que $\Sigma_A \overset{c}{\vee} \Sigma_B$ est un fermé. Donc, sachant que Σ_A et Σ_B sont des fermés, on serait amenés à croire que J_c soit aussi fermé. Or il n'en est rien, comme le démontre le contre-exemple suivant :*

Soit $\Sigma_A = \Sigma_{\{10\}} = \{0^{-\mathbb{N}}1^{\mathbb{N}}, 0^{\mathbb{Z}}, 1^{\mathbb{Z}}\}$ et $\Sigma_B = \Sigma_{\{0\}} = \{1^{\mathbb{Z}}\}$. Alors, $\forall c \in \mathbb{N}, J_c = \{0^{-\mathbb{N}}1^{\mathbb{N}}, 1^{\mathbb{Z}}\}$. On remarque que $0^{\mathbb{Z}} \notin J_c$, donc J_c n'est pas fermé.

Proposition 6. *La c -jointure est associative.*

Cette proposition a déjà été énoncée dans l'article de Enrico Formenti, nous l'avons démontrée d'une autre façon, plus proche de la définition utilisée. La démonstration se trouve dans l'annexe C.

3 Premier résultat sur les langages rationnels

Le premier résultat que nous avons montré est la stabilité des langages rationnels par l'opération de c -jointure.

3.1 Stabilité des langages rationnels par la c -join : cas fini

La définition initiale de la c -join nous a imposé de réaliser la démonstration en exhibant un automate fini. Après avoir remarqué que la c -join pouvait être définie avec $|w| = c$, nous avons pu trouver une expression dénotant la

c -join. La construction de l'automate est tout de même importante pour avoir une vision de la reconnaissabilité de la c -join. Elle suit la démonstration de rationalité.

Théorème 1. Soient L_A et L_B deux langages rationnels de \mathcal{A}^* et $c \in \mathbb{N}$. Alors $L_A \overset{c}{\vee} L_B$ est rationnel.

Démonstration. On a :

$$J_c = \{uvw : w \in \mathcal{A}^c \wedge uw \in \text{Pref}(L_A) \wedge wv \in \text{Suff}(L_B)\} = \bigcup_{w \in \mathcal{A}^c} \text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B)$$

où pour $x \in \mathcal{A}^*$, $Lx^{-1} = \{y \in \mathcal{A}^* : yx \in L\}$ et $x^{-1}L = \{y \in \mathcal{A}^* : xy \in L\}$ sont appelés quotient à droite et quotient à gauche.

Les quotients d'un langage rationnel sont rationnels, et les langages rationnels sont stables par concaténation. J_c est alors une union finie de langages rationnels, et est donc rationnel.

On en déduit que $L_A \overset{c}{\vee} L_B = L_A \cup L_B \cup J_c$ est rationnel. □

Lemme 1. Soit $L_F = \text{Fact}(L_A) \cap \text{Fact}(L_B) \cap \mathcal{A}^c$. Alors

$$L_A \overset{c}{\vee} L_B = L_A \cup L_B \cup \left(\bigcup_{w \in L_F} \text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B) \right)$$

Démonstration.

$$\begin{aligned} L_A \overset{c}{\vee} L_B &= L_A \cup L_B \cup \{uvw : w \in \mathcal{A}^c \wedge uw \in \text{Pref}(L_A) \wedge wv \in \text{Suff}(L_B)\} \\ &= L_A \cup L_B \cup \left(\bigcup_{w \in \mathcal{A}^c} \text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B) \right) \quad (\text{vu à la démonstration précédente}) \\ &= L_A \cup L_B \cup \left(\bigcup_{w \in L_F} \text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B) \right) \end{aligned}$$

Car $\text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B) \neq \emptyset \iff w \in \text{Fact}(L_A) \cap \text{Fact}(L_B)$. □

Construisons l'automate reconnaissant $L_A \overset{c}{\vee} L_B$, dans le cas où L_A et L_B sont non vides.

Soient $A = (\mathcal{A}, Q_A, i_A, F_A, \delta_A)$ et $B = (\mathcal{A}, Q_B, i_B, F_B, \delta_B)$ deux automates finis déterministes émondés tels que $\mathcal{L}(A) = L_A$ et $\mathcal{L}(B) = L_B$.

Soit $F = (\mathcal{A}, Q_F, i_F, F_F, \delta_F)$ un automate fini déterministe tel que $\mathcal{L}(F) = L_F$.

On suppose sans perte de généralité que les états de A , B et F sont tous distincts.

Soit $G = (\mathcal{A} \cup \{\varepsilon\}, Q_G, I_G, F_G, \delta_G)$ un automate fini non déterministe tel que :

- $Q_G = Q_A \cup (Q_A \times Q_F \times Q_B) \cup Q_B$
- $I_G = \{i_A, i_B\}$
- $F_G = F_A \cup F_B$
- $\delta_G = \delta_A \cup \delta_B \cup \delta_P \cup \delta_\varepsilon$

$$\text{où : } \delta_P = \left\{ (q_A q_F q_B, a, q'_A q'_F q'_B) : a \in \mathcal{A} \text{ et } \begin{cases} (q_A, a, q'_A) \in \delta_A \\ (q_F, a, q'_F) \in \delta_F \\ (q_B, a, q'_B) \in \delta_B \end{cases} \right\}$$

qui sont les transitions de l'automate produit $A \times F \times B$

$$\text{et où : } \delta_\varepsilon = \left\{ (q_A, \varepsilon, q_A i_F q_B) : \begin{matrix} q_A \in Q_A \\ q_B \in Q_B \end{matrix} \right\} \cup \left\{ (q_A q_F q_B, \varepsilon, q_B) : \begin{matrix} q_A \in Q_A \\ q_F \in F_F \\ q_B \in Q_B \end{matrix} \right\}$$

La Figure 3 représente l'automate construit dans cette sous-partie. La démonstration de la correction de cette construction se trouve dans l'annexe B.

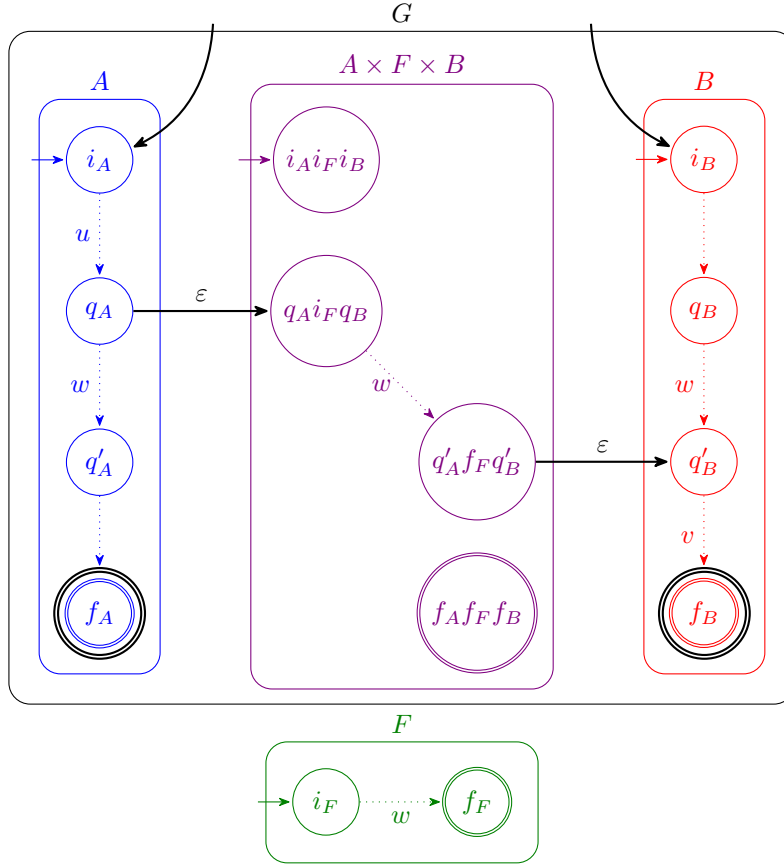


FIGURE 3 – Construction de l'automate G reconnaissant $L_A \overset{c}{\vee} L_B$

3.2 En vue de la généralisation au cas infini

Cette partie définit les automates sur les mots bi-infinis pour donner des idées de la construction de l'automate dans le cas infini.

Voici la définition des automates sur les mots bi-infinis suivante (voir [Car03]) : $A = (\mathcal{A}, Q, I, F, M, \delta)$ où $M \subseteq Q$ est l'ensemble des états milieux. L'ensemble des états Q est toujours fini.

On généralise le terme d'exécution aux suites bi-infinies de transitions $\gamma = (p_i, x_i, q_i)_{i \in \mathbb{Z}} \in \delta^{\mathbb{Z}}$. On a toujours la condition $\forall i \in \mathbb{Z}, q_i = p_{i+1}$. On ajoute la condition $p_0 \in M$. Le mot lu est : $x = \cdots x_{-1} \cdot x_0 x_1 \cdots$.

Une exécution $\gamma \in \delta^{\mathbb{Z}}$ est *initiale* si $\{i \leq 0 : p_i \in I\}$ est infini. Elle est *finale* si $\{i \geq 0 : q_i \in F\}$ est infini. Le mot $x \in \mathcal{A}^{\mathbb{Z}}$ est accepté par l'automate A lorsqu'il existe une exécution initiale et finale dont la lecture est x .

Avec cette définition, la construction précédente devrait se généraliser aux mots bi-infinis. L'automate produit $A \times F \times B$ sera un automate reconnaissant les mots finis, mais l'automate G sera bien un automate sur les mots bi-infinis.

4 Résultats avec les langages facteurs et les mots interdits

On rappelle que les subshifts sont déterminés par leur langage facteur, mais aussi par un ensemble de mots interdits. Nous nous sommes demandés si l'on pouvait caractériser ces ensembles pour les résultats de l'opération de c -jointure sur deux subshifts.

4.1 La c -join et les langages facteurs

Ce théorème permet de déterminer le langage facteur du subshift résultat de la c -join de deux subshifts dont on connaît les langages facteurs.

Théorème 2. Soient Σ_A et Σ_B deux subshifts de $\mathcal{A}^{\mathbb{Z}}$ et $c \in \mathbb{N}$. On a :

$$\text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) = \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)$$

Démonstration. Soit $x \in \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)$.

- Si $x \in \text{Fact}(\Sigma_A)$, il existe $(u, v) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$ tel que $uxv \in \Sigma_A$.
Donc $uxv \in \Sigma_A \overset{c}{\vee} \Sigma_B$, d'où $x \in \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)$.
- Si $x \in \text{Fact}(\Sigma_B)$, idem.
- Sinon $x = u_1 w v_1$ où $w \in \mathcal{A}^c$, $u_1 w \in \text{Pref}(\text{Fact}(\Sigma_A)) = \text{Fact}(\Sigma_A)$ et $w v_1 \in \text{Suff}(\text{Fact}(\Sigma_B)) = \text{Fact}(\Sigma_B)$.
Il existe $(u_2, v_2) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$ tel que $u_2 u_1 w v_1 v_2 \in \Sigma_A$.
Il existe $(v_3, v_4) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$ tel que $v_3 w v_1 v_4 \in \Sigma_B$.
Donc $u_2 x v_4 = u_2 u_1 w v_1 v_4 \in \Sigma_A \overset{c}{\vee} \Sigma_B$ car $w \in \mathcal{A}^c$, $u_2 u_1 w \in \text{Pref}(\Sigma_A)$ et $w v_1 v_4 \in \text{Suff}(\Sigma_B)$.
D'où $x \in \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)$.

Ainsi $\boxed{\text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) \subseteq \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)}$.

Réciproquement, soit $x \in \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)$.

Il existe $(x_1, x_2) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$ tels que $x_1 x x_2 \in \Sigma_A \overset{c}{\vee} \Sigma_B$. Notons $y = x_1 x x_2$.

- Si $y \in \Sigma_A$, alors $x \in \text{Fact}(\Sigma_A) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)$.
- Si $y \in \Sigma_B$, idem.
- Sinon $y = u_1 w v_1$ où $w \in \mathcal{A}^c$, $u_1 w \in \text{Pref}(\Sigma_A)$ et $w v_1 \in \text{Suff}(\Sigma_B)$.
On a $x_1 x x_2 = u_1 w v_1$. Il y a six cas à considérer :

$$\begin{array}{l}
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad x \in \text{Fact}(\Sigma_B) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) \\
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad x \in \text{Fact}(\Sigma_B) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) \\
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad x \in \text{Fact}(\Sigma_A) \cap \text{Fact}(\Sigma_B) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) \\
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad \text{Cas intéressant, développé ci-dessous.} \\
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad x \in \text{Fact}(\Sigma_A) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B) \\
 \text{---} \frac{x_1 \quad | \quad x \quad | \quad x_2}{u \quad | \quad w \quad | \quad v} \text{---} \quad x \in \text{Fact}(\Sigma_A) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)
 \end{array}$$

Cas intéressant :

Il existe $(u_1, u_2) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^*$ et $(v_1, v_2) \in \mathcal{A}^* \times \mathcal{A}^{\mathbb{N}}$ tels que $u = u_1 u_2$, $v = v_1 v_2$ et $x = u_2 w v_1$.

(on remarquera que $x_1 = u_1$ et que $x_2 = v_2$)

On a $u_2 w \in \text{Fact}(\Sigma_A) = \text{Pref}(\text{Fact}(\Sigma_A))$ car $u_1 u_2 w = u w \in \text{Pref}(\Sigma_A)$.

On a $w v_1 \in \text{Fact}(\Sigma_B) = \text{Suff}(\text{Fact}(\Sigma_B))$ car $w v_1 v_2 = w v \in \text{Suff}(\Sigma_B)$.

Et $w \in \mathcal{A}^c$ donc $x = u_2 w v_1 \in \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)$.

Ainsi $\boxed{\text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B) \subseteq \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)}$. □

Ce résultat confirme que la c -join de deux subshifts est un subshift. On peut aussi en déduire une expression de l'ensemble de mots interdits maximal de la c -join :

Corollaire 1.

$$D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) = \overline{\text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)} = \overline{\text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)} = \overline{D^+(\Sigma_A) \overset{c}{\vee} D^+(\Sigma_B)}$$

Ce résultat n'est pas idéal pour travailler avec les ensembles maximaux de mots interdits. Nous avons cherché une autre caractérisation propre aux ensembles de mots interdits. C'est le théorème suivant.

4.2 Caractérisation de l'ensemble maximal des mots interdits

On rappelle que les ensembles de mots interdits ont été définis à la section 1.3.

Nous avons été amenés à utiliser cette définition plus raffinée pour le dernier théorème. Elle permet de décrire les éléments de la c -join n'appartenant à aucun des subshifts initiaux. On appelle ces configurations les *mots inédits* :

$$I_c = \left\{ uawbv : (a, b) \in \mathcal{A}^2, |w| \geq c, \begin{array}{l} uaw \in \text{Pref}(\Sigma_A) \\ uawb \notin \text{Pref}(\Sigma_A) \end{array} \text{ et } \begin{array}{l} wbv \in \text{Suff}(\Sigma_B) \\ awbv \notin \text{Suff}(\Sigma_B) \end{array} \right\}$$

On considère en réalité la plus grande portion de superposition commune (voir figure 4).

La c -join devient donc : $\Sigma_A \overset{c}{\vee} \Sigma_B = \Sigma_A \cup \Sigma_B \cup I_c$

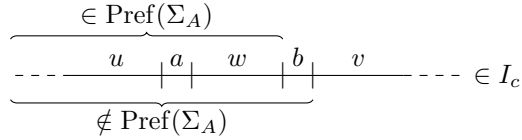


FIGURE 4 – Découpage d'un mot inédit (seules les conditions sur Σ_A ont été indiquées)

Théorème 3. Soit c un entier naturel. Soient deux subshifts Σ_A et Σ_B d'ensembles minimaux de mots interdits D_A^- et D_B^- . Alors pour tout mot fini x :

$$x \in D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) \iff \exists(p_\alpha, d_\alpha)\exists(p_\beta, d_\beta), d_\alpha - p_\beta \leq c \wedge x_{[p_\alpha, d_\alpha]} \in D_A^- \wedge x_{[p_\beta, d_\beta]} \in D_B^-$$

Remarque 2 (Intuition). Un mot x est dans $D^+(\Sigma_A \overset{c}{\vee} \Sigma_B)$ lorsqu'il a pour facteurs deux mots interdits minimaux $\alpha \in D_A^-$ et $\beta \in D_B^-$, tel que β apparaisse après α dans x , ou alors avec une avance d'au plus c lettres sur la dernière lettre de α . Voir la figure 5.

Dans la formule, les entiers $p_\alpha, d_\alpha, p_\beta$ et d_β sont les indices des premières et dernières lettres des mots interdits α et β .

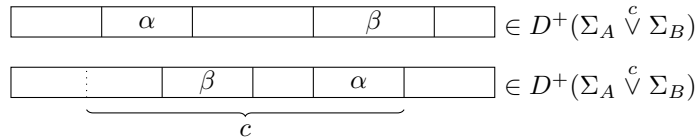


FIGURE 5 – Mots interdits de la c -join

Démonstration. Pour simplifier la notation, commençons par noter :

$$\phi(x) \equiv \exists(p_\alpha, d_\alpha)\exists(p_\beta, d_\beta), d_\alpha - p_\beta \leq c \wedge x_{[p_\alpha, d_\alpha]} \in D_A^- \wedge x_{[p_\beta, d_\beta]} \in D_B^-.$$

On a alors $\neg\phi(x) \equiv \forall(p_\alpha, d_\alpha)\forall(p_\beta, d_\beta), d_\alpha - p_\beta > c \vee x_{[p_\alpha, d_\alpha]} \notin D_A^- \vee x_{[p_\beta, d_\beta]} \notin D_B^-$.

Pour l'implication directe, considérons $x \in D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) = \overline{\text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B)}$ et notons $n = |x|$.

Donc on a $x \in \overline{\text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)} = D_A^+ \cap D_B^+ \cap \overline{W_c}$, d'après le théorème 2, avec

$$W_c = \{uvw : uv \in \text{Fact}(\Sigma_A), vw \in \text{Fact}(\Sigma_B) \text{ et } |w| = c\}.$$

Montrons $\phi(x)$. Soit $(p_\alpha, d_\alpha) \in \llbracket 0, n-1 \rrbracket^2$ tel que $x_{[p_\alpha, d_\alpha]} \in D_A^-$ et tel que $\forall(p'_\alpha, d'_\alpha), x_{[p'_\alpha, d'_\alpha]} \in D_A^- \Rightarrow d_\alpha \leq d'_\alpha$.

Soit $(p_\beta, d_\beta) \in \llbracket 0, n-1 \rrbracket^2$ tel que $x_{[p_\beta, d_\beta]} \in D_B^-$ et tel que $\forall(p'_\beta, d'_\beta), x_{[p'_\beta, d'_\beta]} \in D_B^- \Rightarrow p'_\beta \leq p_\beta$.

De tels couples existent toujours car $x \in D_A^+ \cap D_B^+$. Intuitivement, parmi tous les facteurs de x dans D_A^- , α est celui dont la dernière lettre est la plus à gauche dans x . Idem pour β qui est le facteur de x dont la première lettre est la plus à droite parmi les facteurs de x dans D_B^- . Montrons que $d_\alpha - p_\beta \leq c$.

Supposons par l'absurde que $d_\alpha - p_\beta > c$.

Posons $u = x_{[0, d_\alpha - c - 1]}$, $w = x_{[d_\alpha - c, d_\alpha - 1]}$ et $v = x_{[d_\alpha, n - 1]}$. On a bien $d_\alpha - c > p_\beta \geq 0$ donc $0 \leq d_\alpha - c - 1$ et $d_\alpha \leq n - 1$. Si $c = 0$, on prendra $w = \varepsilon$.

$$x = \underbrace{\boxed{\quad \beta \quad \dots \quad \alpha \quad \dots \quad \quad}}_{u \quad |w|=c \quad v} \in D^+(\Sigma_A \overset{c}{\vee} \Sigma_B)$$

FIGURE 6 – Implication directe, découpage du mot x

Alors, par définition de d_α , $uw \notin D_A^+$ donc $uw \in \text{Fact}(\Sigma_A)$. De même, par définition de $p_\beta < d_\alpha - c$, $vw \notin D_B^+$ donc $vw \in \text{Fact}(\Sigma_B)$. Or $|w| = c$ donc $x = uvw \in W_c$ ce qui est absurde car $x \notin \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)$.

Donc $d_\alpha - p_\beta \leq c$ et de plus $x_{[p_\alpha, d_\alpha]} \in D_A^-$ et $x_{[p_\beta, d_\beta]} \in D_B^-$.

Ainsi on a $\phi(x)$, d'où $\boxed{x \in D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) \implies \phi(x)}$.

Pour la réciproque, nous allons montrer la contraposée : $x \notin D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) \implies \neg\phi(x)$.

Soit $x \notin D^+(\Sigma_A \overset{c}{\vee} \Sigma_B)$. Alors $x \in \text{Fact}(\Sigma_A \overset{c}{\vee} \Sigma_B) = \text{Fact}(\Sigma_A) \overset{c}{\vee} \text{Fact}(\Sigma_B)$ d'après la définition 9 et le théorème 2.

- Si $x \in \text{Fact}(\Sigma_A)$, on a $\forall(p_\alpha, d_\alpha) \in \llbracket 0, |x| - 1 \rrbracket, x_{[p_\alpha, d_\alpha]} \in \text{Fact}(\Sigma_A) = \overline{D_A^+} \subseteq \overline{D_A^-}$ donc $x \notin D_A^-$, d'où $\neg\phi(x)$.
- Si $x \in \text{Fact}(\Sigma_B)$, même raisonnement.
- Sinon $x \in I_c$, $x = uawbv$ tel que $uaw \in \text{Fact}(\Sigma_A)$, $uawb \notin \text{Fact}(\Sigma_A)$, $wbv \in \text{Fact}(\Sigma_B)$, $awbv \notin \text{Fact}(\Sigma_B)$ et $|w| \geq c$. Notons i_a et i_b les indices des lettres a et b dans x . On a en particulier $i_b - i_a = |w| + 1 \geq c + 1$. Soient $p_\alpha, d_\alpha, p_\beta$ et d_β dans $\llbracket 0, |x| - 1 \rrbracket$ tels que $p_\alpha \leq d_\alpha$ et $p_\beta \leq d_\beta$. Montrons $\psi \equiv d_\alpha - p_\beta > c \vee x_{[p_\alpha, d_\alpha]} \notin D_A^- \vee x_{[p_\beta, d_\beta]} \notin D_B^-$.
Si $d_\alpha - p_\beta > c$, on a ψ .
Sinon on a $d_\alpha - p_\beta \leq c$ et on pose $\alpha = x_{[p_\alpha, d_\alpha]}$ et $\beta = x_{[p_\beta, d_\beta]}$.
Si $\alpha \notin D_A^-$, on a ψ .
Sinon $\alpha \in D_A^-$ donc $\alpha \notin \text{Fact}(\Sigma_A)$ donc par la décomposition de x en $uawbv$ on a $i_b \leq d_\alpha$.
Ainsi $i_a + 1 \leq i_b - c \leq d_\alpha - c \leq p_\beta$, donc $\beta \sqsubseteq wbv \in \text{Fact}(\Sigma_B)$.
D'où $\beta \in \text{Fact}(\Sigma_B) = \overline{D_B^+} \subseteq \overline{D_B^-}$ puis ψ .
On a donc montré $\neg\phi(x)$.

D'où $\boxed{x \notin D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) \implies \neg\phi(x)}$ et le sens réciproque.

Ainsi nous avons le résultat : $\boxed{x \in D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) \iff \phi(x)}$

□

Exemple 6. Reprenons l'exemple 5 : $\Sigma_A = \Sigma_{\{111\}}$ et $\Sigma_B = \Sigma_{\{10\}}$. On a $D_A^- = \{111\}$ et $D_B^- = \{10\}$.

Pour $c \leq 3$, on a : $D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) = \mathcal{A}^*111\mathcal{A}^*10\mathcal{A}^* \cup \mathcal{A}^*1110\mathcal{A}^*$.

On en déduit facilement que : $\forall c \leq 3 \ D^-(\Sigma_A \overset{c}{\vee} \Sigma_B) = \{1110\}$.

Pour $c \geq 4$, on a : $D^+(\Sigma_A \overset{c}{\vee} \Sigma_B) = \left(\bigcup_{k \in \llbracket 4, c \rrbracket} \mathcal{A}^*10\mathcal{A}^{c-4}111\mathcal{A}^* \right) \cup \mathcal{A}^*111\mathcal{A}^*10\mathcal{A}^* \cup \mathcal{A}^*1110\mathcal{A}^*$.

On en déduit que : $\forall c \geq 4, \ D^-(\Sigma_A \overset{c}{\vee} \Sigma_B) = \{10^{k-3}111 \mid k \in \llbracket 4, c \rrbracket\} \cup \{1110\}$.

Exemple avec $c = 5$: $D^-(\Sigma_A \overset{5}{\vee} \Sigma_B) = \{100111, 10111, 1110\}$. On remarque que l'on a ajouté des mots interdits.

Conclusion et perspectives

Les résultats obtenus sont assez variés, mais les derniers théorèmes portent leur intérêt sur les caractérisations des subshifts. Le travail sur les subshifts est très orienté sur leurs ensembles de mots interdits, le dernier théorème permet de se passer de la description des subshifts par leurs configurations et de se contenter des ensembles de mots interdits.

Il serait intéressant de regarder les effets de la c -jointure sur les langages qui engendrent un subshift. Nous ne nous sommes pas non plus penchés sur la question de la distributivité de la c -join sur d'autres opérations sur les langages ou les subshifts.

On pourrait aussi se demander comment calculer efficacement l'ensemble $D^-(\Sigma)$ d'un subshift Σ à partir d'un autre ensemble de mots interdits de Σ . On a réalisé pendant le stage des manipulations sur ces ensembles, mais rien n'est encore formalisé.

Ce stage m'a permis de découvrir le travail de chercheur en laboratoire de recherche. J'ai pu me familiariser avec les notions de dynamique symbolique étudiées (subshifts, topologie sur les mots bi-infinis).

Références

- [BH85] François BLANCHARD et Georges HANSEL : Languages and subshifts. *In Automata on Infinite Words*, pages 137–146, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg.
- [BP97] Marie-Pierre BÉAL et Dominique PERRIN : *Symbolic Dynamics and Finite Automata*, pages 463–506. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [Buz05] Jérôme BUZZI : Subshifts of quasi-finite type. *Inventiones mathematicae*, 159(2):369–406, 2005.
- [Car03] Olivier CARTON : Unambiguous automata on bi-infinite words. *In Mathematical Foundations of Computer Science 2003*, pages 308–317, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.
- [FK07] Enrico FORMENTI et Petr KŮRKA : A search algorithm for the maximal attractor of a cellular automaton. *In STACS 2007*, pages 356–366, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [Kŭr03] Petr KŮRKA : *Topological and symbolic dynamics*, volume 11 de *Cours spécialisés*, chapitre 3. Société mathématique de France, 2003.
- [LM95] Douglas LIND et Brian MARCUS : *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
- [Pav22] Ronnie PAVLOV : On subshifts with slow forbidden word growth. *Ergodic Theory and Dynamical Systems*, 42(4):1487–1516, 2022.
- [PP04] Dominique PERRIN et Jean-Éric PIN : *Infinite Words*. Pure and Applied Mathematics Vol 141. Elsevier, 2004.
- [Pro12] Julien PROVILLARD : *Automates cellulaires non-uniformes*. Thèse de doctorat, 2012.

Annexes

A Définition complète des préfixes, suffixes et facteurs

Voici comment définir les préfixes, suffixes et facteurs d'un mot de manière combinatoire.

| | Pref(x) | Suff(x) |
|-----------------------------------|--|--|
| $x \in \mathcal{A}^*$ | $\{u \in \mathcal{A}^* : \exists v \in \mathcal{A}^*, uv = x\}$ | $\{v \in \mathcal{A}^* : \exists u \in \mathcal{A}^*, uv = x\}$ |
| $x \in \mathcal{A}^{\mathbb{N}}$ | $\{u \in \mathcal{A}^* : \exists v \in \mathcal{A}^{\mathbb{N}}, uv = x\}$ | $\{v \in \mathcal{A}^{\mathbb{N}} : \exists u \in \mathcal{A}^*, uv = x\}$ |
| $x \in \mathcal{A}^{-\mathbb{N}}$ | $\{u \in \mathcal{A}^{-\mathbb{N}} : \exists v \in \mathcal{A}^*, uv = x\}$ | $\{v \in \mathcal{A}^* : \exists u \in \mathcal{A}^{-\mathbb{N}}, uv = x\}$ |
| $x \in \mathcal{A}^{\mathbb{Z}}$ | $\{u \in \mathcal{A}^{-\mathbb{N}} : \exists v \in \mathcal{A}^{\mathbb{N}} \exists n \in \mathbb{Z}, \sigma^n(u \cdot v) = x\}$ | $\{v \in \mathcal{A}^{\mathbb{N}} : \exists u \in \mathcal{A}^{-\mathbb{N}} \exists n \in \mathbb{Z}, \sigma^n(u \cdot v) = x\}$ |
| $x \in \mathcal{A}^{\zeta}$ | $\{u \in \mathcal{A}^{-\mathbb{N}} : \exists v \in \mathcal{A}^{\mathbb{N}}, uv = x\}$ | $\{v \in \mathcal{A}^{\mathbb{N}} : \exists u \in \mathcal{A}^{-\mathbb{N}}, uv = x\}$ |

| | Fact(x) |
|-----------------------------------|--|
| $x \in \mathcal{A}^*$ | $\{w \in \mathcal{A}^* : \exists (u, v) \in (\mathcal{A}^*)^2, u w v = x\}$ |
| $x \in \mathcal{A}^{\mathbb{N}}$ | $\{w \in \mathcal{A}^* : \exists (u, v) \in \mathcal{A}^* \times \mathcal{A}^{\mathbb{N}}, u w v = x\}$ |
| $x \in \mathcal{A}^{-\mathbb{N}}$ | $\{w \in \mathcal{A}^* : \exists (u, v) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^*, u w v = x\}$ |
| $x \in \mathcal{A}^{\mathbb{Z}}$ | $\{w \in \mathcal{A}^* : \exists (u, v) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}} \exists n \in \mathbb{Z}, \sigma^n((u \cdot w) \cdot v) = x\}$ |
| $x \in \mathcal{A}^{\zeta}$ | $\{w \in \mathcal{A}^* : \exists (u, v) \in \mathcal{A}^{-\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}, u w v = x\}$ |

B Correction de la construction de l'automate

Démonstration. Montrons que $\mathcal{L}(G) = L_A \overset{c}{\vee} L_B$.

Soit $x \in L_A \overset{c}{\vee} L_B$.

- Si $x \in L_A$, alors il existe $f_A \in F_A$ tel que $i_A \xrightarrow{A} f_A$.
Or $i_A \in I_G$, $f_A \in F_G$ et $\delta_A \subseteq \delta_G$.
Donc $x \in \mathcal{L}(G)$.
- Si $x \in L_B$, même raisonnement.
- Sinon $x = u w v$ où $u \in \text{Pref}(L_A) w^{-1}$, $w \in L_F$ et $v \in w^{-1} \text{Suff}(L_B)$.
Soit $(q_B, q'_B, f_B) \in Q_B \times Q_B \times F_B$ tel que $q_B \xrightarrow{B} q'_B \xrightarrow{B} f_B$, qui existe car $w v \in \text{Suff}(L_B)$.
Soit $q_F \in Q_F$ tel que $i_F \xrightarrow{F} q_F$, qui existe car $w \in L_F$.
Soit $(q_A, q'_A) \in Q_A \times Q_A$ tel que $i_A \xrightarrow{A} q_A \xrightarrow{A} q'_A$, qui existe car $u w \in \text{Pref}(L_A)$.
Alors on a :

$$i_A \xrightarrow{G} q_A \xrightarrow{G} q_A i_F q_B \xrightarrow{G} q'_A q_F q'_B \xrightarrow{G} q'_B \xrightarrow{G} f_B$$

et $i_A \in I_G$, $f_B \in F_G$ donc $x = u w v \in \mathcal{L}(G)$.

D'où $L_A \overset{c}{\vee} L_B \subseteq \mathcal{L}(G)$.

Réciproquement, soit $x \in \mathcal{L}(G)$.

Soient $(q_0, \dots, q_n) \in Q_G^{n+1}$ et $(x_1, \dots, x_n) \in (\mathcal{A} \cup \{\varepsilon\})^n$ tels que :

- $x = x_1 x_2 \cdots x_n$
- $(q_0, q_n) \in I_G \times F_G$
- $q_0 \xrightarrow{G} q_1 \xrightarrow{G} q_2 \xrightarrow{G} \cdots \xrightarrow{G} q_{n-1} \xrightarrow{G} q_n$.

La construction de l'automate G assure que seuls trois cas sont à considérer :

- Si $q_0 = i_A$ et $q_n \in F_A$, alors $\forall m \in \llbracket 0, n \rrbracket$, $q_i \in Q_A$ donc $x \in L_A \subseteq L_A \overset{c}{\vee} L_B$.

- Si $q_0 = i_B$ et $q_n \in F_B$, même chose.
- Sinon $q_0 = i_A$ et $q_n \in F_B$.

Par construction, il existe un unique couple d'entiers (k, l) tels que $1 \leq k < l \leq n$ et $x_k = x_l = \varepsilon$.

Alors $\forall m \in \llbracket 0, k-1 \rrbracket, q_m \in Q_A, \forall m \in \llbracket k, l-1 \rrbracket, q_m = (q_A^m q_F^m q_B^m) \in Q_A \times Q_F \times Q_B$ et $\forall m \in \llbracket l, n \rrbracket, q_m \in Q_B$.

Notons $u = x_1 \cdots x_{k-1}, u = x_{k+1} \cdots x_{l-1}$ et $v = x_{l+1} \cdots x_n$.

Donc : $q_0 \xrightarrow[A]{u} q_{k-1} \xrightarrow[G]{\varepsilon} q_k \xrightarrow[G]{w} q_{l-1} \xrightarrow[G]{\varepsilon} q_l \xrightarrow[B]{v} q_n$

On a : $q_A^k \xrightarrow[A]{w} q_A^{l-1}, q_F^k \xrightarrow[F]{w} q_F^{l-1}$ et $q_B^k \xrightarrow[B]{w} q_B^{l-1}$

et $q_A^k = q_{k-1}, q_F^k = i_F, q_F^{l-1} = F_F$ et $q_B^{l-1} = q_l$.

Alors $w \in L_F$.

On a $i_A = q_0 \xrightarrow[A]{u} q_{k-1} = q_A^k \xrightarrow[A]{w} q_A^{l-1}$ donc $uw \in \text{Pref}(L_A)$ car tout état est accessible dans A .

Alors $u \in \text{Pref}(L_A)w^{-1}$.

On a $q_B^k \xrightarrow[B]{w} q_B^{l-1} = q_l \xrightarrow[B]{v} q_n \in F_B$ donc $wv \in \text{Suff}(L_B)$ car tout état est co-accessible dans B .

Alors $v \in w^{-1} \text{Suff}(L_B)$.

Ainsi $x = u\varepsilon w\varepsilon v = uwv \in \text{Pref}(L_A)w^{-1} \cdot w \cdot w^{-1} \text{Suff}(L_B)$ et $w \in L_F$.

Donc $x \in L_A \overset{c}{\vee} L_B$.

D'où $\mathcal{L}(G) \subseteq L_A \overset{c}{\vee} L_B$.

Ainsi $\mathcal{L}(G) = L_A \overset{c}{\vee} L_B$ donc $L_A \overset{c}{\vee} L_B$ est reconnaissable donc rationnel. □

C Démonstration de l'associativité

Preuve de la proposition 6. Démontrons le résultat sur les subshifts. Pour les langages de mots finis, la démonstration est identique.

Soient Σ_A, Σ_B et Σ_C trois subshifts de $\mathcal{A}^{\mathbb{Z}}$. Montrons que $(\Sigma_A \overset{c}{\vee} \Sigma_B) \overset{c}{\vee} \Sigma_C = \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

Soit $x \in (\Sigma_A \overset{c}{\vee} \Sigma_B) \overset{c}{\vee} \Sigma_C$. Plusieurs cas peuvent se présenter :

- Si $x \in \Sigma_A \overset{c}{\vee} \Sigma_B$ alors il faudra examiner trois possibilités supplémentaires :

- Si $x \in \Sigma_A$, alors $x \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$

- Si $x \in \Sigma_B$, alors $x \in \Sigma_B \overset{c}{\vee} \Sigma_C \subseteq \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

- Sinon $x = uvw$ où $w \in \mathcal{A}^c, uw \in \text{Pref}(\Sigma_A)$ et $wv \in \text{Suff}(\Sigma_B)$.

Soit v' tel que $v'wv \in \Sigma_B$. Alors $v'wv \in \Sigma_B \overset{c}{\vee} \Sigma_C$.

Donc $wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$.

Donc $x = uvw \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

- Si $x \in \Sigma_C$, alors $x \in \Sigma_B \overset{c}{\vee} \Sigma_C \subseteq \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

- Sinon $x = uvw$ où $w \in \mathcal{A}^c, uw \in \text{Pref}(\Sigma_A \overset{c}{\vee} \Sigma_B)$ et $wv \in \text{Suff}(\Sigma_C)$.

Soit $(u', v') \in \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{-\mathbb{N}}$ tel que $uwu' \in \Sigma_A \overset{c}{\vee} \Sigma_B$ et $v'wv \in \Sigma_C \subseteq \Sigma_B \overset{c}{\vee} \Sigma_C$.

Montrons que $x \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

Posons $y = uwu' \in \Sigma_A \overset{c}{\vee} \Sigma_B$.

- Si $y \in \Sigma_A$, alors $uw \in \text{Pref}(\Sigma_A)$.

Et $wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C), w \in \mathcal{A}^c$.

D'où $x = uvw \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

- Si $y \in \Sigma_B$, alors $uw \in \text{Pref}(\Sigma_B)$.

Et $wv \in \text{Suff}(\Sigma_C), w \in \mathcal{A}^c$.

- D'où $x = uvw \in \Sigma_B \overset{c}{\vee} \Sigma_C \subseteq \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.
- Sinon $y = u_1w_1v_1$ où $w_1 \in \mathcal{A}^c$, $u_1w_1 \in \text{Pref}(\Sigma_A)$ et $w_1v_1 \in \text{Suff}(\Sigma_B)$.
Soit $(u'_1, v'_1) \in \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{-\mathbb{N}}$ tel que $u_1wu'_1 \in \Sigma_A$ et $v'_1wv_1 \in \Sigma_B$.

$$\begin{array}{c}
 x \text{ ----- } \begin{array}{|c|c|c|} \hline u & w & v \\ \hline \end{array} \text{ -----} \\
 = \\
 \begin{array}{c}
 \left\{ \begin{array}{l}
 \text{----- } \begin{array}{|c|c|c|} \hline u & w & u' \\ \hline u_1 & w_1 & v_1 \\ \hline \end{array} \text{-----} \text{ cas 1} \\
 \text{----- } \begin{array}{|c|c|c|} \hline u & w & u' \\ \hline u_1 & & v_1 \\ \hline \end{array} \text{-----} \text{ cas 2} \\
 \text{----- } \begin{array}{|c|c|c|} \hline u & w & u' \\ \hline u_1 & w_1 & v_1 \\ \hline \end{array} \text{-----} \text{ cas 3} \\
 \text{----- } \begin{array}{|c|c|c|} \hline u & w & u' \\ \hline u_1 & & v_1 \\ \hline \end{array} \text{-----} \text{ cas 4} \\
 \text{----- } \begin{array}{|c|c|c|} \hline u & w & u' \\ \hline & w_1 & v_1 \\ \hline u_1 & & \\ \hline \end{array} \text{-----} \text{ cas 5}
 \end{array}
 \right.
 \end{array}$$

1. On a $x = uvw = u_1w_1\beta wv$.
On a $v'_1w_1\beta w \in \text{Pref}(\Sigma_B)$ car $\beta w \in \text{Pref}(v_1)$ et $v'_1w_1v_1 \in \Sigma_B$. De plus $wv \in \text{Suff}(\Sigma_C)$ et $w \in \mathcal{A}^c$.
Donc $w_1\beta wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$.
On a alors : $u_1w_1 \in \text{Pref}(\Sigma_A)$, $w_1\beta wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$ et $w_1 \in \mathcal{A}^c$.
D'où $x = u_1w_1\beta wv \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.
2. On a $x = uvw = u_1\alpha wv_1 = u_1w_1\beta v$.
On a $v'_1\alpha w \in \text{Pref}(\Sigma_B)$ car $v'_1\alpha w = v'_1w_1\beta$, $\beta \in \text{Pref}(v_1)$ et $v'_1w_1v_1 \in \Sigma_B$. De plus $wv \in \text{Suff}(\Sigma_C)$ et $w \in \mathcal{A}^c$.
Donc $w_1\beta v \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$.
On a alors : $u_1w_1 \in \text{Pref}(\Sigma_A)$, $w_1\beta v \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$ et $w_1 \in \mathcal{A}^c$.
D'où $x = u_1w_1\beta v \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.
3. On a $u = u_1$ et $w = w_1$.
On a : $uw = u_1w_1 \in \text{Pref}(\Sigma_A)$, $wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$ et $w \in \mathcal{A}^c$.
D'où $x = uvw \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.
4. On a : $u' = \alpha v_1$.
On a : $uw \in \text{Pref}(\Sigma_A)$ car $uw\alpha = u_1w_1 \in \text{Pref}(\Sigma_A)$, $wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$ et $w \in \mathcal{A}^c$.
D'où $x = uvw \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.
5. On a $u' = \alpha w_1v_1$.
On a : $uw \in \text{Pref}(\Sigma_A)$ car $uw\alpha = u_1w_1 \in \text{Pref}(\Sigma_A)$, $wv \in \text{Suff}(\Sigma_B \overset{c}{\vee} \Sigma_C)$ et $w \in \mathcal{A}^c$.
D'où $x = uvw \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$.

D'où $x \in \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)$

Ainsi $\boxed{(\Sigma_A \overset{c}{\vee} \Sigma_B) \overset{c}{\vee} \Sigma_C \subseteq \Sigma_A \overset{c}{\vee} (\Sigma_B \overset{c}{\vee} \Sigma_C)}$.

La preuve de l'autre inclusion est totalement symétrique, d'où le résultat. \square